


```
Matches 124; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Qy 89 CTCCTGACACCTATNTGCATCAAGTGCCTCAATGACCAATGACAGTGTGCGGTTAGACAAC 148
Db 63680 CTTTAAACAACCTTCTGAGCATGGTGCAATGACAAGGATAGTGTAGAAGTACACAAC 63621
Qy 149 TACTATACGGGCAAGGAAGTGGAGATTGAGTTGGATGGCTTTGACTCTCTAGCCAAAT 208
Db 63620 TACTATACCTGGGAAAGATTACCAATCCACTAAATGTTGGCTCTAACGCCCTAACCAAGAT 63561
Qy 209 GCCAGCGGTACTTCAAGAAGTACCAGAACTCAAGGAGGGGCTCAAGCACTGA 263
Db 63560 GCTCAGCGTTACTTCAAGAAATACCAGAAGCTCAAGGAAGCTGTTAAGCACTTAA 63506

RESULT 7
AR120267
LOCUS AR120267 1423 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 59 from patent US 6159469.
ACCESSION AR120267
VERSION AR120267.1 GI:14103843
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 1423)
AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B., Fannon, M.R. and Rosen, C.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6159469-A 59 12-DEC-2000;
FEATURES
source
Location/Qualifiers
1..1423
/mol_type="unassigned DNA"

ORIGIN
Query Match 32.3%; Score 85; DB 6; Length 1423;
Best Local Similarity 67.8%; Pred. No. 1.9e-15; Indels 0; Gaps 0;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
Qy 90 TCCTGACCACCTATNTGCATCAAGTGCCTCAATGACCAATGACAGTGTGCGGTTAGACAAC 149
Db 717 TGCTGACAACCTTCTCCACCAAGTGCCTAAGCAAGACCAAGCTTATCCTAGACAAC 776
Qy 150 ACTATACGGGCAAGGAAGTGGAGATTGAGTTGGATGGCTTTGACTCTCTAGCCAAATG 209
Db 777 ACTATACCAACCACTATCATGATTGCGTTGATTAAGGCTCTGACTCCCAACCAAGATG 836
Qy 210 CCCAGCGGTACTTCAAGAAGTACCAGAACTCAAGGAGGGGCTCAAGCACTGA 263
Db 837 CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAAGCTGTCAAAATACTTGA 890

RESULT 9
AR340958
LOCUS AR340958 1423 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 59 from patent US 6573082.
ACCESSION AR340958
VERSION AR340958.1 GI:33732937
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 1423)
AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B., Fannon, M.R. and Rosen, C.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6573082-A 59 03-JUN-2003;
FEATURES
source
Location/Qualifiers
1..1423
/mol_type="genomic DNA"

ORIGIN
Query Match 32.3%; Score 85; DB 6; Length 1423;
Best Local Similarity 67.8%; Pred. No. 1.9e-15; Indels 0; Gaps 0;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
Qy 90 TCCTGACCACCTATNTGCATCAAGTGCCTCAATGACCAATGACAGTGTGCGGTTAGACAAC 149
Db 717 TGCTGACAACCTTCTCCACCAAGTGCCTAAGCAAGACCAAGCTTATCCTAGACAAC 776
Qy 150 ACTATACGGGCAAGGAAGTGGAGATTGAGTTGGATGGCTTTGACTCTCTAGCCAAATG 209
Db 777 ACTATACCAACCACTATCATGATTGCGTTGATTAAGGCTCTGACTCCCAACCAAGATG 836
Qy 210 CCCAGCGGTACTTCAAGAAGTACCAGAACTCAAGGAGGGGCTCAAGCACTGA 263
Db 837 CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAAGCTGTCAAAATACTTGA 890

RESULT 10
AR653221
LOCUS AR653221 1423 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 59 from patent US 6987663.
```

```
CL2N15/31, C12N5/18, C12N1/21, C07K14/315, C12Q1/68, A61K39/09, PC
G01N33/569;
PC G01N33/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
Location/Qualifiers
1..1423
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Query Match 32.3%; Score 85; DB 6; Length 1423;
Best Local Similarity 67.8%; Pred. No. 1.9e-15; Indels 0; Gaps 0;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
Qy 90 TCCTGACCACCTATNTGCATCAAGTGCCTCAATGACCAATGACAGTGTGCGGTTAGACAAC 149
Db 717 TGCTGACAACCTTCTCCACCAAGTGCCTAAGCAAGACCAAGCTTATCCTAGACAAC 776
Qy 150 ACTATACGGGCAAGGAAGTGGAGATTGAGTTGGATGGCTTTGACTCTCTAGCCAAATG 209
Db 777 ACTATACCAACCACTATCATGATTGCGTTGATTAAGGCTCTGACTCCCAACCAAGATG 836
Qy 210 CCCAGCGGTACTTCAAGAAGTACCAGAACTCAAGGAGGGGCTCAAGCACTGA 263
Db 837 CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAAGCTGTCAAAATACTTGA 890

RESULT 9
AR340958
LOCUS AR340958 1423 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 59 from patent US 6573082.
ACCESSION AR340958
VERSION AR340958.1 GI:33732937
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 1423)
AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B., Fannon, M.R. and Rosen, C.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6573082-A 59 03-JUN-2003;
FEATURES
source
Location/Qualifiers
1..1423
/mol_type="genomic DNA"

ORIGIN
Query Match 32.3%; Score 85; DB 6; Length 1423;
Best Local Similarity 67.8%; Pred. No. 1.9e-15; Indels 0; Gaps 0;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
Qy 90 TCCTGACCACCTATNTGCATCAAGTGCCTCAATGACCAATGACAGTGTGCGGTTAGACAAC 149
Db 717 TGCTGACAACCTTCTCCACCAAGTGCCTAAGCAAGACCAAGCTTATCCTAGACAAC 776
Qy 150 ACTATACGGGCAAGGAAGTGGAGATTGAGTTGGATGGCTTTGACTCTCTAGCCAAATG 209
Db 777 ACTATACCAACCACTATCATGATTGCGTTGATTAAGGCTCTGACTCCCAACCAAGATG 836
Qy 210 CCCAGCGGTACTTCAAGAAGTACCAGAACTCAAGGAGGGGCTCAAGCACTGA 263
Db 837 CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAAGCTGTCAAAATACTTGA 890

RESULT 10
AR653221
LOCUS AR653221 1423 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 59 from patent US 6987663.
```

RESULT 12	AR481615	1683 bp	DNA	linear	PAT 14-MAY-2004
LOCUS	Sequence 2519 from patent US 6699703.				
DEFINITION	AR481615				
ACCESSION	AR481615				
VERSION	AR481615.1	GI:47240577			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 1683)				
AUTHORS	Doucette-Stamm,L., Bush,D., Zeng,Q., Opperman,T. and Houseweart,C.E.				
TITLE	Nucleic acid and amino acid sequences relating to Streptococcus pneumoniae for diagnostics and therapeutics				
JOURNAL	Patent: US 6699703-A 2519 02-MAR-2004;				
FEATURES	Genome Therapeutics Corporation; Waltham, MA				
source	Location/Qualifiers				
	1..1683				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN					
Query Match	32.3%;	Score 85;	DB 6;	Length 1683;	
Best Local Similarity	67.8%;	Pred. No. 1.9e-15;			
Matches 118;	Conservative 0;	Mismatches 56;	Indels 0;	Gaps 0;	
QY	90 TCCTGACCAACCTATNTGCATCAAGTGCACAAATGACCAAGTCCAGTGTGCGGTTAGACAACT 149				
Db	974 TGCTGACAACTTCTCTCCACCAAGTGCCTAACGACCAAGACCAAGTTCCTAGACAACT 1033				
QY	150 ACTATACGGGCAAGAACTGAGATTGAGTTGGATGTGGCTTTGACTCTCTAGCCAAATG 209				
Db	1034 ACTATACCAACCAACCTATCATGATTGCGCTTGATTAAGGCTCTGACTCCCAACCAAGATG 1093				
QY	210 CCCAGCGGTACTTCAAGAGTACAGAACTCAAGGAGGCGGTCAAGCAACCTGA 263				
Db	1094 CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAAAGCTGTCAATACTTTGA 1147				
RESULT 13	AR587857	1692 bp <td>DNA</td> <td>linear</td> <td>PAT 15-DEC-2004</td>	DNA	linear	PAT 15-DEC-2004
LOCUS	Sequence 1976 from patent US 6800744.				
DEFINITION	AR587857				
ACCESSION	AR587857				
VERSION	AR587857.1	GI:56632114			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 1692)				
AUTHORS	Doucette-Stamm,L.A. and Bush,D.				
TITLE	Nucleic acid and amino acid sequences relating to Streptococcus pneumoniae for diagnostics and therapeutics				
JOURNAL	Patent: US 6800744-A 1976 05-OCT-2004;				
FEATURES	Genome Therapeutics Corporation; Waltham, MA				
source	Location/Qualifiers				
	1..1692				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN					
Query Match	32.3%;	Score 85;	DB 6;	Length 1692;	
Best Local Similarity	67.8%;	Pred. No. 1.9e-15;			
Matches 118;	Conservative 0;	Mismatches 56;	Indels 0;	Gaps 0;	
QY	90 TCCTGACCAACCTATNTGCATCAAGTGCACAAATGACCAAGTCCAGTGTGCGGTTAGACAACT 149				
Db	983 TGCTGACAACTTCTCTCCACCAAGTGCCTAACGACCAAGACCAAGTTCCTAGACAACT 1041				
QY	150 ACTATACGGGCAAGAACTGGAGATTGATTGGATGTGGCTTTTGACTCTTAGTCCCAAAATG 209				

Db 1043 ACTATACCAACCACTATCATGTTGGCTTGATAGGCTCTGACTCCCAACAGATG 1102
Qy 210 CCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGGAGCGGTCAAGCACTGA 263
Db 1103 CCCAAGCGTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAATACTTGA 1156

RESULT 14
AF181976 2015 bp DNA linear BCT 02-NOV-1999
LOCUS Streptococcus pneumoniae adherence and virulence protein A (pva)
DEFINITION gene, complete cds.
ACCESSION AF181976
VERSION AF181976.1 GI:6175914
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE 1 (bases 1 to 2015)
Holmes, A.R., McNab, R., Millsap, K. and Jenkinson, H.F.
TITLE The pva gene of Streptococcus pneumoniae encodes a
fibronectin-binding protein that is necessary for Pneumococcal cell
adhesion and virulence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2015)
Holmes, A.R., McNab, R. and Jenkinson, H.F.
AUTHORS Direct Submission
TITLE Submitted (31-AUG-1999) Oral Sciences and Orthodontics, University
JOURNAL of Otago, Great King St, Dunedin, New Zealand
LOCATION/Qualifiers

FEATURES
source
1..2015
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/strain="R800"
/db_xref="taxon:1313"
98..1763
/gene="pva"
98..102
/gene="pva"
108..1763
/gene="pva"
/function="adhesin"
/codon_start=1
/transl_table=11
/product="adherence and virulence protein A"
/protein_id="AA05332.1"
/db_xref="GI:6175915"
/translation="MSFPGFLHIVELRSELVNGRIQKINQPEQLVLQIRSNRQ
SHRLLLSHPVGRILQTLTTPENPAQSTFIMVLKYLQGLALIESIQVENDRIVEI
TVSNKNEIGDIQATLIIEINGKHSNILLVDKSHKILEVHKHVFQSONSYRTLLPGS
TYIAPPSTENPPTIKDEKLEIQLQELTAKNLQSLFQGLGRDTANELERILVSEK
LSAFNPNQKCTCLTFTSPVFNQAGEPANLSDLDTYVKNKAERDRVKQQA
SELIRRVENELKQKHKLQKQERELLADNAEERQKGLLTTLHQVPNDQDVIDL
NYTNPQIMIALDKALTNQNAQRYFKYQKLEKAVKYLTLIETKATILYLSVET
VLNQGLSEIAIEIRELLIQTFIRRRQREKIQKRKKUEQYLASDGKTIIVYGRNNLQN
BELTFMKARKELMFAKDIPIGSHVVISGNLDPDSDAVKTDAAELAAVPSQGRLSNLVQ
VDMIEVKLNKPTGGKPGFVTVTGQKTLRVTPDPSKKIASMKKS"

ORIGIN

Query Match 32.3%; Score 85; DB 1; Length 2015;
Best Local Similarity 67.8%; Pred. No. 1.9e-15;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
Qy 90 TCCTGACCACCTATNTGATCAAGTCCAAATGACCACTCGAGTGTGGGTAGCAACT 149
Db 1054 TGCTGACCACTTCTCCACCAAGTCCCTACGACCAAGACCAAGCTTATCCTAGCAACT 1113
Qy 150 ACTATACCGGCAAGGAACTGGAGATGATGGATGTGGCTTTGACTCCTAGCCAAATG 209
Db 1114 ACTATACCAACCACTATCATGATTTGGCTTGATAGGCTCTGACTCCCAACCAAGATG 1173

Qy 210 CCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGGAGCGGTCAAGCACTGA 263
Db 1174 CCCAAGCGTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAATACTTGA 1227

RESULT 15
CQ788941/c 10240 bp DNA linear PAT 29-MAR-2004
LOCUS Sequence 32 from Patent EP1400592.
DEFINITION CQ788941
ACCESSION CQ788941
VERSION CQ788941.1 GI:45822509
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1
unclassified sequences.

REFERENCE 1
Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Barash, S.C.,
Fannon, M. and Dougherty, B.A.
TITLE Streptococcus pneumoniae polynucleotides and sequences
JOURNAL Patent: EP 1400592-A 32 24-MAR-2004;
HUMAN GENOME SCIENCES, INC. (US)
FEATURES
source
1..10240
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

ORIGIN

Query Match 32.3%; Score 85; DB 6; Length 10240;
Best Local Similarity 67.8%; Pred. No. 2e-15;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
Qy 90 TCCTGACCACCTATNTGATCAAGTCCAAATGACCACTCGAGTGTGGGTAGCAACT 149
Db 4646 TGCTGACCACTTCTCCACCAAGTCCCTACGACCAAGACCAAGCTTATCCTAGCAACT 4587
Qy 150 ACTATACCGGCAAGGAACTGGAGATTTGAGTTGGATGTGGCTTCTAGTCCCAAAATG 209
Db 4586 ACTATACCAACCACTATCATGATTTGGCTTGATAGGCTCTGACTCCCAACCAAGATG 4527
Qy 210 CCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGGAGCGGTCAAGCACTGA 263
Db 4526 CCCAAGCGTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAATACTTGA 4473

Search completed: January 20, 2006, 15:56:27
Job time : 2163 secs

CC may be used in the diagnosis of and vaccination against streptococcal
CC infections and in the detection of virulence markers of Streptococci. The
CC vaccine comprising the clone, nucleic acid, vector, or the host cell, is
CC also useful in the prevention and/or treatment of streptococcal
CC infections. ABS59787-ABS59832 represent Streptococcus suis in vivo
CC selected (i)ve) genes and related PCR primers of the invention

XX Sequence 263 BP; 72 A; 63 C; 61 G; 66 T; 0 U; 1 Other;

SQ Query Match 99.0%; Score 260.4; DB 6; Length 263;

Best Local Similarity 99.6%; Pred. No. 9e-79;

Matches 262; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ACGAAATDGTGATGATCGATCATGATAAAGTGGATGCTTAACTGTTTTCGTGCTAT 60

Db 1 ACGAAATCGATGATGATCGATCATGATAAAGTGGATGCTTAACTGTTTTCGTGCTAT 60

Qy 61 TTTTGTGAATCGAATTCGAGCTCGCCCTCTGACCACTTATNTGCAATCAAGTGCATA 120

Db 61 TTTTGTGAATCGAATTCGAGCTCGCCCTCTGACCACTTATNTGCAATCAAGTGCATA 120

Qy 121 TGACCAAGTCGAGTGTGCGGTTAGACAACTACTATACCGGCAAGAACTGGAGATTGAGTT 180

Db 121 TGACCAAGTCGAGTGTGCGGTTAGACAACTACTATACCGGCAAGAACTGGAGATTGAGTT 180

Qy 181 GGATGTGGCTTTGACTCTAGCCAAATGCCAGGGTACTTCAAGAGTACCAAGAACT 240

Db 181 GGATGTGGCTTTGACTCTAGCCAAATGCCAGGGTACTTCAAGAGTACCAAGAACT 240

Qy 241 CAAGAGGCGGTCAAGCACTGA 263

Db 241 CAAGAGGCGGTCAAGCACTGA 263

RESULT 2

AAV27353

ID AAV27353 standard; DNA; 1423 BP.

XX AAV27353;

AC AAV27353;

DT 02-OCT-1998 (first entry)

XX Streptococcus pneumoniae SP0039 nucleotide.

DE Streptococcus pneumoniae.

XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;

KW detection; pneumonia; otitis media; meningitis; ss.

KW Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

OS Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

CC	pneumoniae infection. The present sequence is that of an S pneumoniae
CC	gene of the invention.
XX	
QQ	Sequence 1653 BP; 551 A; 383 C; 322 G; 397 T; 0 U; 0 Other;
	Query Match 32.3%; Score 85; DB 12; Length 1653;
	Best Local Similarity 67.8%; Pred. No. 3.6e-16;
	Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY	90 TCCTGACCACTATNTGATCAAGTGCCAAATGACCAGTCGAGTGTGCGGTTAGACAACT 149
DB	947 TGTGACAACTTCTCTCCACCACTATCATGATTGGCTTTGATAAGGCTCTGACTCCCAACCAAGATG 1066
QY	150 ACTATACGGGCAAGGAACCTGAGATGTGAGTTGGCTTTGACTCTCTAGCCAAATG 209
DB	1007 ACTATACCAACCACTTCTCTCCACCACTATCATGATTGGCTTTGATAAGGCTCTGACTCCCAACCAAGATG 1066
QY	210 CCCAGCGGTACTTCAAGAAGTACAGAACTCAAGGAGGGGCTCAAGCACTTGA 263
DB	1067 CCCAAGCTATTTTAAACGGGTATCAGAACTCAAGAAGGCTGTCAAAATACTTGA 1120
RESULT 6	
ABX06664	
ID	ABX06664 standard; DNA; 1680 BP.
XX	
XX	ABX06664;
XX	
DT	27-OCT-2003 (revised)
DT	11-FEB-2003 (first entry)
XX	
XX	S. pneumoniae type 4 strain coding region #952.
XX	
KW	Gene; ds: bacterial meningitis; pneumonia; sepsis; otitis media;
KW	ear infection; antinflammatory; antibacterial; immunostimulant;
KW	auditory; respiratory; gene therapy; vaccine.
XX	
OS	Streptococcus pneumoniae; type 4 strain.
XX	
PN	WO200277021-A2.
XX	
PD	03-OCT-2002.
XX	
PF	27-MAR-2002; 2002WO-IB002163.
XX	
XX	27-MAR-2001; 2001GB-00007658.
PR	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
XX	
PI	Masignani V, Tettelin H, Fraser C;
XX	
DR	WPI; 2003-040579/03.
DR	P-PSDB; ABU01377.
XX	
PT	New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT	useful as medicaments for treating or preventing a disease or infection
PT	due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT	ear infection.
XX	
PS	Claim 6; SEQ ID NO 1903; 56pp; English.
XX	
CC	The invention relates to a protein comprising or having at least 50%
CC	identity to any of the 2469 amino acid sequences, identified in the
CC	specification (available on a computer readable format), or its fragment,
CC	expressed from 2469 of 2489 identified DNA coding regions from the
CC	Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC	ABS56454. Also included are an antibody which binds one of the proteins,
CC	treating a patient by administering the protein, DNA or antibody (in a
CC	composition), a kit comprising first and second primers, which are the
CC	nucleic acid cited above or fragments between nucleotides 8-100 of a
CC	sequence not defined in the specification, for amplifying a target
CC	sequence contained within a Streptococcus nucleic acid sequence, where


```
DE Streptococcus pneumoniae gene, Seq ID No 2519.
XX ds; gene; Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
KW Streptococcus pneumoniae.
XX US6699703-B1.
XX 02-MAR-2004.
XX 26-MAY-2000; 2000US-00583110.
XX 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
PI WPI; 2004-212399/20.
XX P-PSDB; ADK48665.
XX New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
XX screening.
XX Disclosure; SEQ ID NO 2519; 301pp; English.
XX The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing, preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae nucleic acid of the invention. Note: The
CC sequence data for this patent did not appear in the printed specification
CC but was obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
XX Sequence 1683 BP; 563 A; 386 C; 332 G; 402 T; 0 U; 0 Other;
Query Match 32.3%; Score 85; DB 13; Length 1683;
Best Local Similarity 67.8%; Pred. No. 3.6e-18;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCCTGACCACCTATNTGTCATCAAGTGCACCAATGACACGTGCGGTAGACAACT 149
DB 974 TGCTGACAACTTCTCCACCAAGTGCTTACGACCAAGACGAGTTATCTAGACAACT 1033
QY 150 ACTATACGGGCAAGAACTGGAGATTGAGTTGGATTGGCTTTGACTCTTAGCCAAATG 209
DB 1034 ACTATACCAACCAACTATCATGATTGGCTTGATAGGCTCTGACTCCCAACCAATG 1093
QY 210 CCCAGCGTACTTCAAGAGTACCAAGAACTCAAGGAGCGGTCAAGCACCTGA 263
DB 1094 CCCAACGCTATTTAAACGGTATCAGAACTCAAGAAAGCTGTCAATACTTTGA 1147
RESULT 10
AD93341
ID ADR93341 standard; DNA; 1692 BP.
XX AC ADR93341;
XX 16-DEC-2004 (first entry)
XX Novel S. pneumoniae DNA sequence, SEQ ID 1976.
XX Meningitis; bacteraemia; pneumonia; otitis media; ds;
KW bacterial infection.
XX Streptococcus pneumoniae.
OS
```

```
XX US6800744-B1.
XX 05-OCT-2004.
XX 30-JUN-1998; 98US-00107433.
XX 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm LA, Bush D;
PI WPI; 2004-697205/68.
XX P-PSDB; ADR95944.
XX New isolated nucleic acid encoding a Streptococcus pneumoniae
PT polypeptide, useful for diagnosing, preventing and/or treating
PT pathological conditions resulting from the bacterial infection.
XX Disclosure; SEQ ID NO 1976; 151pp; English.
XX The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridisable under high stringency conditions to the nucleotide sequence.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae nucleic acid sequences. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
XX Sequence 1692 BP; 567 A; 387 C; 332 G; 406 T; 0 U; 0 Other;
Query Match 32.3%; Score 85; DB 13; Length 1692;
Best Local Similarity 67.8%; Pred. No. 3.6e-18;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCCTGACCACCTATNTGTCATCAAGTGCACCAATGACACGTGCGGTAGACAACT 149
DB 983 TGCTGACAACTTCTCCACCAAGTGCTTACGACCAAGACGAGTTATCTAGACAACT 1042
QY 150 ACTATACGGGCAAGAACTGGAGATTGAGTTGGATTGGCTTTGACTCTTAGCCAAATG 209
DB 1043 ACTATACCAACCAACTATCATGATTGGCTTGATAGGCTCTGACTCCCAACCAATG 1102
QY 210 CCCAGCGTACTTCAAGAGTACCAAGAACTCAAGGAGCGGTCAAGCACCTGA 263
DB 1103 CCCAACGCTATTTAAACGGTATCAGAACTCAAGAAAGCTGTCAATACTTTGA 1156
RESULT 11
AEA57211
ID AEA57211 standard; DNA; 1692 BP.
XX AC AEA57211;
XX 25-AUG-2005 (first entry)
DT
```

XX Streptococcus pneumoniae ORF nucleic acid sequence SEQ ID NO:1976.
DE bacterial infection; Streptococcus pneumoniae infection; antibacterial;
KW vaccine; gene; ds.
XX Streptococcus pneumoniae.
OS US2005136404-A1.
XX 23-JUN-2005.
XX 10-JUL-2003; 2003US-00617320.
XX 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX (DOUC/) DOUCETTE-STAMM L A.
PA (BUSH/) BUSH D.
XX Doucette-Stamm LA, Bush D;
PI WPI; 2005-477576/48.
XX P-PSDB; AEA59814.
DR
XX New isolated nucleic acid molecules and encoded polypeptides useful for
PT diagnosing, preventing or treating bacterial infections, particularly
PT Streptococcus pneumoniae infection.
XX
PS Claim 1; SEQ ID NO 1976; 144pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule for detecting,
CC preventing or treating pathological conditions resulting from bacterial
CC infection. The isolated nucleic acid comprises: (a) any of the 2603
CC nucleotide sequences of AEA5236 to AEA57838; (b) a nucleotide sequence
CC encoding a Streptococcus pneumoniae polypeptide comprising any of the
CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide
CC sequence of at least 8 nucleotides in length, where the sequence is
CC hybridizable to a nucleic acid having any of the nucleotide sequences in
CC (a). Also described: (1) a recombinant expression vector comprising the
CC above nucleic acid operably linked to a transcription regulatory element;
CC (2) a cell comprising the recombinant expression vector; (3) producing an
CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence
CC consisting of at least 8 nucleotides of any of AEA5236 to AEA57838; (5)
CC treating a subject for S. pneumoniae infection; (6) a recombinant or its
CC substantially pure preparation of an S. pneumoniae polypeptide or its
CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;
CC (7) a vaccine composition for preventing or treating an S. pneumoniae
CC infection, comprising an amount of the above nucleic acid or polypeptide;
CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;
CC (9) a computer readable medium having recorded the nucleotide sequences
CC of AEA5236 to AEA57838; (10) a computer based system for identifying
CC fragments of the Streptococcus genome of commercial importance. The
CC composition and methods are useful for diagnosing, preventing or treating
CC bacterial infections, particularly S. pneumoniae infection. The present
CC sequence represents a S. pneumoniae ORF nucleic acid sequence from the
CC present invention. Note - The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from the USPTO web site.
XX
XX Sequence 1692 BP; 567 A; 387 C; 332 G; 406 T; 0 U; 0 Other;
SQ
Query Match 32.3%; Score 85; DB 14; Length 1692;
Best Local Similarity 67.8%; Pred. No. 3.6e-18;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCTGTACCACTATNTGCATCAAGTCCCAATGACCAAGTGCAGTGTGGGTAGACAACT 149
DB 983 TGTGTGCAACCTTCTCTCCACCAAGTGCCTAACGACCAAGACCAAGGTATCTTAGACAACT 1042
QY 150 ACTATACGGCAAGGAACTGAGATTGAGTTGGATGTGGCTTTTGACTCTCTAGCCAAATG 209

DB 1043 ACTATACCAACCAACCTATCATGATTTGGCTTGATAGGCTTCTGATCCCAACCAAGT 1102
QY 210 CCCAGCGGTACTTCAAGAAAGTACCAGAACTCAAGGAGCGGTCAAGCACCTGA 263
DB 1103 CCCACGCTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAAAATACTTGA 1156
RESULT 12
AAV52165/c
ID AAV52165 standard; DNA; 10240 BP.
XX AAV52165;
AC AAV52165;
XX 23-OCT-1998 (first entry)
DT Streptococcus pneumoniae genome fragment SEQ ID NO:32.
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
DE computer readable medium; vaccine; pharmaceutical composition; ds.
KW Streptococcus pneumoniae.
OS WO9818931-A2.
XX PN
XX 07-MAY-1998.
PD
XX 30-OCT-1997; 97WO-US019588.
PF
XX 31-OCT-1996; 96US-0029960P.
PR (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
PI Dougherty BA;
PI WPI; 1998-272225/24.
DR
XX Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
XX
PS Claim 1; Page 326-332; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has the
CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
CC 391, identifying members of the library which contain sequences that
CC hybridize to the target sequence and isolating the nucleic acid molecules
CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an
CC organism, amplifying nucleic acid molecules whose nucleotide sequence is
CC homologous to amplification primers derived from the fragment of the S.
CC pneumoniae genome to prime the amplification and isolating the amplified
CC sequences. The computer readable medium can be used in a computer-based
CC system for identifying fragments of the S. pneumoniae genome of
CC commercial importance, or expression modulating fragments of the S.
CC pneumoniae genome. Products from the present invention can be used in
CC diagnosis kits and assays, and pharmaceutical compositions and vaccines
CC for S. pneumoniae
XX
SQ Sequence 10240 BP; 2865 A; 1914 C; 2390 G; 3068 T; 0 U; 3 Other;
Query Match 32.3%; Score 85; DB 2; Length 10240;
Best Local Similarity 67.8%; Pred. No. 8.1e-18;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCTGTACCACTATNTGCATCAAGTCCCAATGACCAAGTGCAGTGTGGGTAGACAACT 149

(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 1653 BP; 548 A; 360 C; 306 G; 439 T; 0 U; 0 Other;

Query Match	31.6%	Score 83.2	DB 8	Length 1653
Best Local Similarity	68.0%	Pred. NO. 1.5e-17		
Matches 115	Conservative 0	Mismatches 54	Indels 0	Gaps 0
QY	93	TGACCACCTATNTGCAATCAAGTGCCTAAATGACCACTCGAGTGTGCGGTTAGACAACTACT	152	
Db	947	TGACCACCTACCTCAGTCTGGTCCCAACACCAAGACTCTGTGATTTAGACAATATT	1006	
QY	153	ATACGGGGCAAGGAAGTGGAGATTGAGTTGGATGTGGCTTTGACTCTTAGCCAAATGCCCC	212	
Db	1007	ATACAGGGGAAAAAATTGAGATTGCCTTAGACAAGGCTCTGACCAAAATCAAAATGCTC	1066	
QY	213	AGCGGTACTTCAAGAGTACCAAGAACTCAAGGAGGGGTCAAGCACT	261	
Db	1067	AACGTTATTTTAAAAAATACCAAAAGCTAAAGAGCTGTCAACACTT	1115	

Search completed: January 20, 2006, 15:20:15
Job time : 475 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2006, 14:03:25 ; Search time 3837 seconds
(without alignments)
3206.933 Million cell updates/sec

Title: US-10-632-117-37
Perfect score: 263
Sequence: 1 acgaataatdgtgatccat.....ggaggcgtcaagcacctga 263

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues
Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	14.4	428	5	BY453989
2	37.6	14.3	807	6	CB898978
3	37.4	14.2	1101	10	CN900L12
4	37	14.1	850	5	BX368104
5	36.8	14.0	406	7	CV072680
6	36.8	14.0	648	5	CA036082
7	36.2	13.8	427	3	BM638966
8	36.2	13.8	476	3	BM642048
9	36.2	13.8	486	3	BM636853
10	36.2	13.8	578	3	BM616582
11	36.2	13.8	585	3	BM624215
12	36.2	13.8	607	3	BM628044
13	36.2	13.8	610	3	BM625696
14	36.2	13.8	615	3	BM578413
15	36.2	13.8	630	3	BM618626
16	36.2	13.8	645	3	BM642729
17	36.2	13.8	648	3	BM627773
18	36.2	13.8	649	3	BM649206
19	36.2	13.8	661	3	BM648844
20	36.2	13.8	670	3	BM632022
21	36.2	13.8	674	3	BM636042
22	36.2	13.8	678	3	BM636340

23	36.2	13.8	683	3	BM651776
24	36.2	13.8	684	3	BM625442
25	36.2	13.8	691	3	BM643948
26	36.2	13.8	692	3	BM653819
27	36.2	13.8	693	3	BM656762
28	36.2	13.8	697	3	BM631562
29	36.2	13.8	709	3	BM592722
30	36.2	13.8	711	3	BM641312
31	36.2	13.8	714	3	BM646862
32	36.2	13.8	715	3	BM627009
33	36	13.7	307	10	CG618897
34	36	13.7	629	5	BM648199
35	35.8	13.6	614	3	BM686992
36	35.6	13.5	441	9	AZ108235
37	35.6	13.5	653	6	CA449674
38	35.6	13.5	713	2	BE970931
39	35.4	13.5	177	1	AW553179
40	35.4	13.5	289	2	BG146742
41	35.4	13.5	299	1	AA086663
42	35.4	13.5	323	10	CG887081
43	35.4	13.5	365	2	BE197026
44	35.4	13.5	500	1	AA689925
45	35.4	13.5	545	6	CA322313

ALIGNMENTS

RESULT 1
BY453989
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BY453989 RIKEN full-length enriched, pooled tissues, adult spleen, etc. Mus musculus cDNA clone K630070L23 3', mRNA sequence.
BY453989.1 GI:26750723
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 428)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Kikaido, I., Oatso, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Bruscia, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. I., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Negashima, T., Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ring, B. Z., Ringwald, M., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Reid, R. Z., Shimada, K., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tonita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakakume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Azawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
PUBMED 12466851

COMMENT

Contact: Yoshihide Hayaahizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa.K., Akimura.T., Arakawa.T., Carninci.P., Fukuda.S., Hirozane.T., Imotani.K., Ishii.Y., Itoh.M., Kawai.J., Konno.H., Miyazaki.A., Murata.M., Nakamura.M., Nomura.K., Numazaki.R., Ohno.M., Sakai.K., Sakazume.N., Sasaki.D., Sato.K., Shibata.K., Shiraki.T., Tagami.M., Waki.K., Watahiki.A., Muramatsu.M. and Hayashizaki.Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES

source Location/Qualifiers

```

1..428
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="K630070L23"
/clone_lib="RIKEN full-length enriched, pooled tissues, adult spleen, etc."
/notes="(dev stage=adult, tissue type=spleen, sex=male), (dev stage=adult, tissue type=kidney, sex=male), (dev stage=adult, tissue type=testis, sex=male), (dev stage=adult, tissue type=thymus, sex=male), (dev stage=adult, tissue type=heart, sex=male), (dev stage=adult, tissue type=colon, sex=male), (dev stage=adult, tissue type=stomach, sex=male), (dev stage=adult, tissue type=liver, sex=male), (dev stage=13 days embryo, tissue type=whole body, sex=mix), (dev stage=14 days embryo, tissue type=whole body, sex=mix), (dev stage=16 days embryo, tissue type=whole body, sex=mix), (dev stage=17 days embryo, tissue type=whole body, sex=mix), (dev stage=15 days pregnant, adult, tissue type=amion, sex=female), (dev stage=10 days neonate, tissue type=brain, sex=mix), (dev stage=10 days neonate, tissue type=thymus, sex=mix), (dev stage=10 days neonate, tissue type=heart, sex=mix)"

```

ORIGIN

```

Query Match      14.4%; Score 38; DB 5; Length 428;
Best Local Similarity 54.8%; Pred. No. 0.87;
Matches 74; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 87 CCCTCTGACCACTATATGTCATCAAGTGCACAAATGACCAAGTCGAGTGTGGGTTAGACA 146
Db 255 CCGCGCGGCGCAGCTGTAGCTGTATTGCTCGGACCCGCGCGCTGTGGCGGAGCG 314
QY 147 ACTACTATACGGCAAGGAACCTGGAGATTGAGTTGACTTGTGCTTGTACTCTAGCCAAA 206
Db 315 ACGGCTATGTGTCGAAGCCAGGGGCTGTGATTGTATCTGCGAAGATCCAAGCCCCGA 374
QY 207 ATGCCAGCGGTACT 221

```

Db 375 AAGGCCAATGAGCCT 389

RESULT 2

CB898978

LOCUS

DEFINITION

CB898978 807 bp mRNA linear EST 02-JUL-2003

tricol16xm10 T.reesei mycelial culture, Version 3 april Hypocrea

jecorina cDNA clone tricol16xm10, mRNA sequence.

ACCESSION

CB898978

VERSION

CB898978.1 GI:30113636

KEYWORDS

EST.

SOURCE

Hypocrea jecorina (anamorph: Trichoderma reesei)

ORGANISM

Hypocrea jecorina

REFERENCE

1 (bases 1 to 807)

Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,

Dunn-Coleman,N.S., Goedegebuur,P., Houfek,T.D., England,G.J.,

Kelley,A.S., Mearman,H.J., Mitchell,T., Mitchinson,C.,

Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.

filamentous fungus Trichoderma reesei

J. Biol. Chem. 278 (34), 31988-31997 (2003)

CONTACT: Pamela K. Foreman

Genencor Intl.

925 Page Mill Road, Palo Alto, CA 94304, USA

Tel: (650) 846-7635

Fax: (650) 621-7817

Email: Pforeman@genencor.com

Seq primer: LT-F1 primer.

Location/Qualifiers

1..807

/organism="Hypocrea jecorina"

/mol_type="mRNA"

/strain="QM6a"

/db_xref="taxon:51453"

/clones="tricol16xm10"

/clone_lib="mycelia"

/clone_lib="T.reesei mycelial culture, Version 3 april"

/note="Vector: pREP3; Site 1: Not I/Sal I; Mycelial

culture grown from 24 hrs to 6 days with varying Carbon

and Nitrogen sources and concentrations."

ORIGIN

Query Match 14.3%; Score 37.6; DB 6; Length 807;

Best Local Similarity 40.8%; Pred. No. 1.4;

Matches 80; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 61 TTTTGTGATCGAATCGAGCTGCCCTCTGACCACCTATATGTCATCAAGTGCCAAA 120

Db 1 TTTTGTGATCGAATCGAATCGAATCGAATCGAATCGAATCGAATCGAATCGAATCG 60

QY 121 TGACCAAGTCGAGTGTGGGTTAGACAACTACTATACGGGCAAGGAACCTGGAGATTGAGTT 180

Db 61 CGTCGGCAGGTTCTCAAGGAGCGCAAGCTCAAGAGCGAGGTTGACGACATGCTTCT 120

QY 181 GGATGTGGCTTTGACTCTTAGCCAAAATGCCAGCGGTPACTTCAAGAGTACCAAGAACT 240

Db 121 GGTGGGGGTTTCCACCCCGTATCCCAAGGTTTCAGTCTCTTATCGAGGAGTACTTTAAGCG 180

QY 241 CAAGGAGCGGTCAAG 256

Db 181 CAAGAAGGCTTCCAAG 196

RESULT 3

CBNS00112

LOCUS

DEFINITION

CBNS00112 1101 bp DNA linear GSS 03-JUN-1999

Drosophila melanogaster genome survey sequence TERT end of BAC:

BACR23D10 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION

AL067834

```
VERSION AL067834.1 GI:4957863
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
        Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
        Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
        Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
        BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
        - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
        collaboration with the Berkeley Drosophila Genome Project (BDGP).
        The BDGP is constructing a physical map of the Drosophila
        melanogaster genome using these BACs. For further information
        please see http://www.fruitfly.org The BDGP Drosophila
        melanogaster BAC library was prepared by Kazutoyo Osoegawa and
        Aaron Mammosser in Pieter de Jong's laboratory in the Department of
        Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
        NY. The library is named Rpci-98 and was constructed by partial
        EcoRI digestion of Drosophila DNA provided by the BDGP from the
        isogenic strain y2; cn bw 9p, the same strain used for the BDGP's
        p1 and EST libraries. A more detailed description of the library
        and how to order individual BAC clones, the entire library, or
        filters for hybridization from the BACPAC Resource Center can be
        found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             source
   Location/Qualifiers
       1..1101
           /organism="Drosophila melanogaster"
           /mol_type="genomic DNA"
           /db_xref="taxon:7227"
           /clone="BACR23D10"
           /clone_lib="RPCI-98"
           /note="end : TET3"

ORIGIN
Query Match      14.2%; Score 37.4; DB 10; Length 1101;
Best Local Similarity 33.2%; Pred. No. 1.7;
Matches 68; Conservative 47; Mismatches 90; Indels 0; Gaps 0;

QY 58 TATTTTTCGTGAATCGAATCGAGCTCGCCCTCTCGACACCTATNTGCATCAAGTGCC 117
Db 1051 TATTTTTCGTGAATCGAATCGAATCGAGCTCGCCCTCTCGACACCTATNTGCATCAAGTGCC 117

QY 118 AAATGACCACTGAGTGTGGGTAGACAACACTATATACGGCAAGAACTGGAGATTGA 177
Db 991 AAAAAAAAWGGRGGVADDDGRAMDMAVCCHAAMSVVAAAMTMCCAAAGVSGSGR 932

QY 178 GTTGGATGTGGTTTGCATCTCTAGCAAAATGCCAGCGGTACTTCAAGAAGTACCAGAA 237
Db 931 DNCGGGGGGGAAGGSDTCCCBKCAACACGCGSVGAGRWDDGGRRRAAWTDAAGG 872

QY 238 ACTCAGAGGCGGTCAAGCACTG 262
Db 871 GGGRRAAARGGGRRDRAWAAACKG 847

RESULT 4
LOCUS BX368104/c
DEFINITION BX368104 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
        clone CS0D1073YC17 3-PRIME, mRNA sequence.
ACCESSION BX368104
VERSION BX368104
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
        Homnidae; Homo.
REFERENCE 1 (bases 1 to 850)

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
        Genoscope - Centre National de Sequencage
        2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
        Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
        1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
        end enriched, double-strand cDNA was digested with Not I and cloned
        into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
        was normalized. Library was constructed by Life Technologies, a
        division of Invitrogen. This sequence belongs to sequence cluster
        7464.f
        For more information about this cluster, see
        http://www.genoscope.cns.fr/cdna?e=CS0AU013ZD11_U01236_l&c=7464.f.

FEATURES             source
   Location/Qualifiers
       1..850
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /clone="CS0D1073YC17"
           /clone_lib="PLACENTA COT 25-NORMALIZED"
           /note="1st strand cDNA was primed with a NotI-oligo(dT)
           primer. Five prime end enriched, double-strand cDNA was
           digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match      14.1%; Score 37; DB 5; Length 850;
Best Local Similarity 54.3%; Pred. No. 2.2;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 135 TCGCGTTAGACAACACTATATACGGCAAGGAAGTCTGAGATTGAGTTGGCTTTGA 194
Db 775 TCGCGTACCGCCAGCTGNNACAAAGGCCAGATCCAGAGATCGTCTGCTGGTGGCGGTCCA 716

QY 195 CTCCTAGCCAAATGCCAGCGGTACTTCAAGAAGTACCAGAACTCAAGAGCGCGTCA 254
Db 715 CTCGTATCCCAAGATCCAGAGCTGCTCAGAGATTCTTCAACGCAAGGAGCTCAACA 656

QY 255 AGCAGCTGA 263
Db 655 AGAGCATCA 647

RESULT 5
LOCUS CV072680
DEFINITION Lr JV2CF 27G04 SKplus Juvenile Earthworm Library Lumbricus rubellus
        cDNA clone Lr JV2CF_27G04, mRNA sequence.
ACCESSION CV072680
VERSION CV072680
KEYWORDS EST.
SOURCE Lumbricus rubellus (humus earthworm)
ORGANISM Lumbricus rubellus
        Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;
        Lumbricina; Lumbricidae; Lumbricus.
REFERENCE 1 (bases 1 to 406)
AUTHORS Chaeley, J., Hedley, B.A., Morgan, A.J., Sturzenbaum, S., Kille, P. and
        Blaxter, M.
TITLE The Lumbricus rubellus EST program - Sequences from a Juvenile
        library
JOURNAL Unpublished (2004)
COMMENT Contact: Jennifer Chaeley
        BIOS1
        Cardiff University
        Main College, Museum Avenue, Cardiff, CF11 3TL, UK
        Tel: +44 2920876680
        Fax: +44 2920874305
        Email: chaeley@cardiff.ac.uk, Kille@cardiff.ac.uk
        Sequencing was performed in Edinburgh using the pBluescript II XR
        cDNA library (Stratagene) protocol.
```

```

PCR Primers
FORWARD: M13P (gtttccagtcacgacg)
BACKWARD: M13R (caggaacagctatgaccatg)
Plate: 27 row: G column: 04
Seq primer: SKplus
High quality sequence start: 8
High quality sequence stop: 351.
FEATURES
    source
        1..406
        /organism="Lumbricus rubellus"
        /mol_type="mRNA"
        /db_xref="taxon:35632"
        /clone="Lr_JV2CF_27G04"
        /sex="mixed"
        /tissue_type="Whole worm"
        /dev_stage="Juvenile"
        /clone_lib="Juvenile Earthworm Library"
        /note="Organ: Whole worm"
ORIGIN
Query Match          14.0%; Score 36.8; DB 7; Length 406;
Best Local Similarity 66.2%; Pred. No. 2.1;
Matches 53; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Qy 160 CAAGAACTGGAGATTGAGTTGGATGGCTTTGACTCTAGCCAAAATGCCAGCGTA 219
Db 97 CAACAACTGGAGTTGCTTGGATGGAGCGAACAAGAGCGCGCAGAGCGCGAGAAGAA 156
Qy 220 CTTCAAGAAGTACCAGAAAC 239
Db 157 CATCAGAAGTTCAGCAAC 176
RESULT 6
CA036082
LOCUS
DEFINITION
Lr_ade_01G12_T3 Earthworm Lambda Zap Express Library Lumbricus
rubellus cDNA clone Lr_ade_01G12 5' similar to pir|A59287 myosin
heavy chain - fluke (Schistosoma mansoni) (strain Brazilian, mRNA
sequence.
ACCESSION
CA036082
VERSION
CA036082.1 GI:24334779
KEYWORDS
EST.
SOURCE
Lumbricus rubellus (humus earthworm)
ORGANISM
Lumbricus rubellus
Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;
Lumbricina; Lumbricidae; Lumbricidae.
REFERENCE
1 (bases 1 to 648)
AUTHORS
Sturzenbaum,S., Parkinson,J., Blaxter,M., Morgan,J., Kille,P.,
Schaffner,W. and Georgiev,O.
Expressed Sequence Tags from the humus earthworm L. rubellus
Unpublished (2000)
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared using protocol given by supplier
(Stratagene).
PCR Primers
FORWARD: T3
BACKWARD: T7PL
Plate: 01 row: G column: 12
Seq primer: T3
High quality sequence stop: 457.
FEATURES
    source
        1..648
        /organism="Lumbricus rubellus"
        /mol_type="mRNA"
        /db_xref="taxon:35632"

```

```

/clone="Lr_ade_01G12"
/tissue_type="Whole worm"
/dev_stage="adult"
/clone_lib="Earthworm Lambda Zap Express Library"
/note="Vector: pBK-CMV; The library was prepared using
protocol given by supplier (Stratagene)."
ORIGIN
Query Match          14.0%; Score 36.8; DB 5; Length 648;
Best Local Similarity 66.2%; Pred. No. 2.4;
Matches 53; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Qy 160 CAAGAACTGGAGATTGAGTTGGATGGCTTTGACTCTAGCCAAAATGCCAGCGTA 219
Db 275 CAAGAACTGGAGTTGCTTGGATGGAGCGAACAAGAGCGCGCAGAGCGCGAGAAGAA 334
Qy 220 CTTCAAGAAGTACCAGAAAC 239
Db 335 CATCAGAAGTTCAGCAAC 354
RESULT 7
BM638966
LOCUS
DEFINITION
17000687567125 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
19600449621565 5', mRNA sequence.
ACCESSION
BM638966
VERSION
BM638966.1 GI:18938477
KEYWORDS
EST.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE
1 (bases 1 to 427)
AUTHORS
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltR@celera.com
Plate: NU01004N8J row: E column: 19
Seq primer: M13 Reverse.
FEATURES
    source
        1..427
        /organism="Anopheles gambiae"
        /mol_type="mRNA"
        /strain="RSP-ST (Reduced susc. to Permethrin - std.
        chromosome)"
        /db_xref="taxon:7165"
        /clone="19600449621565"
        /dev_stage="Adult"
        /lab_host="DHI10b"
        /clone_lib="A.Gam.ad.cDNA1"
        /note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
        adult mosquitoes (mixed sex)-frozen on liquid nitrogen.
        cDNA inserts >500 bp cloned directionally into pSport 1.
        Not 1 site is 3'. Clones available through the Malaria
        Research and Reference Reagent Resource Center
        (www.malaria.mr4.org)."
ORIGIN
Query Match          13.8%; Score 36.2; DB 3; Length 427;
Best Local Similarity 60.8%; Pred. No. 3.3;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
Qy 160 CAAGAACTGGAGATTGAGTTGGATGGCTTTGACTCTAGCCAAAATGCCAGCGTA 219
Db 211 CAACGAGCTGGAGATTGCTCTGGATCAGCGCAACAAGGCTAACGCTGAGGCCAGAGAA 270

```

```

QY      220  CTTCAAGAGTACCAAGAACTCAAGGAGCGGTCAAG 256
DB      271  CATCAAGCGCTACCAGCAGCAGCTGAAGGACGTCCAG 307

RESULT 8
BM642048
LOCUS   17000687309375 A.Gam.ad.cdNA1 Anopheles gambiae cDNA clone
DEFINITION
ACCESSION   BM642048
VERSION     19600449651236 5', mRNA sequence.
KEYWORDS    EST.
SOURCE      BM642048.1 GI:18941559
ORGANISM    Anopheles gambiae (African malaria mosquito)
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Anophelinae; Anopheles.
            1 (bases 1 to 476)
            Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,
            Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
            Celera Anopheles gambiae EST project
            Unpublished (2002)
            Contact: Holt R.A.
            Celera Genomics
            45 W. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU010049W6 row: J column: 02
            Seq primer: M13 Reverse
            Location/Qualifiers
            1..476
            /organism="Anopheles gambiae"
            /mol_type="mRNA"
            /strain="RSP-ST (Reduced susc. to Permethrin - std.
            chromosome)"
            /db_xref="taxon:7165"
            /clone="19600449651236"
            /dev_stage="Adult"
            /lab_host="DH10b"
            /clone_lib="A.Gam.ad.cdNA1"
            /note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
            adult mosquitoes (mixed sex) frozen on liquid nitrogen.
            cDNA inserts >500 bp cloned directionally into pSport 1.
            Not 1 site is 3'. Clones available through the Malaria
            Research and Reference Reagent Resource Center
            (www.malaria.mr4.org)."

FEATURES
            source
            1..476
            /organism="Anopheles gambiae"
            /mol_type="mRNA"
            /strain="RSP-ST (Reduced susc. to Permethrin - std.
            chromosome)"
            /db_xref="taxon:7165"
            /clone="19600449651236"
            /dev_stage="Adult"
            /lab_host="DH10b"
            /clone_lib="A.Gam.ad.cdNA1"
            /note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
            adult mosquitoes (mixed sex) frozen on liquid nitrogen.
            cDNA inserts >500 bp cloned directionally into pSport 1.
            Not 1 site is 3'. Clones available through the Malaria
            Research and Reference Reagent Resource Center
            (www.malaria.mr4.org)."

ORIGIN
            Query Match      13.8%; Score 36.2; DB 3; Length 476;
            Best Local Similarity 60.8%; Pred. No. 3.4;
            Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY      160  CAAAGAACTGGAGATTGAGTTGGATGGCTTGTGACTCTAGCCAAATGCCAGCGTA 219
DB      249  CAACGAGCTGGAGATTGCTCTGGATCAGCCCAACAGGCTAACGCTGAGGCCCAAGA 308

RESULT 9
BM636853
LOCUS   17000687562979 A.Gam.ad.cdNA1 Anopheles gambiae cDNA clone
DEFINITION
ACCESSION   BM636853
VERSION     19600449642127 5', mRNA sequence.
KEYWORDS    EST.
SOURCE      BM636853.1 GI:18936364
ORGANISM    Anopheles gambiae (African malaria mosquito)
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Anophelinae; Anopheles.
            1 (bases 1 to 486)
            Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,
            Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
            Celera Anopheles gambiae EST project
            Unpublished (2002)
            Contact: Holt R.A.
            Celera Genomics
            45 W. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com

```

```

ORGANISM    Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Anophelinae; Anopheles.
            1 (bases 1 to 486)
            Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,
            Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
            Celera Anopheles gambiae EST project
            Unpublished (2002)
            Contact: Holt R.A.
            Celera Genomics
            45 W. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01003CV2 row: N column: 13
            Seq primer: M13 Reverse
            Location/Qualifiers
            1..486
            /organism="Anopheles gambiae"
            /mol_type="mRNA"
            /strain="RSP-ST (Reduced susc. to Permethrin - std.
            chromosome)"
            /db_xref="taxon:7165"
            /clone="19600449642127"
            /dev_stage="Adult"
            /lab_host="DH10b"
            /clone_lib="A.Gam.ad.cdNA1"
            /note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
            adult mosquitoes (mixed sex) frozen on liquid nitrogen.
            cDNA inserts >500 bp cloned directionally into pSport 1.
            Not 1 site is 3'. Clones available through the Malaria
            Research and Reference Reagent Resource Center
            (www.malaria.mr4.org)."

ORIGIN
            Query Match      13.8%; Score 36.2; DB 3; Length 486;
            Best Local Similarity 60.8%; Pred. No. 3.4;
            Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY      160  CAAAGAACTGGAGATTGAGTTGGATGGCTTGTGACTCTAGCCAAATGCCAGCGTA 219
DB      277  CAACGAGCTGGAGATTGCTCTGGATCAGCCCAACAGGCTAACGCTGAGGCCCAAGA 336

RESULT 10
BM616582
LOCUS   17000687148935 A.Gam.ad.cdNA1 Anopheles gambiae cDNA clone
DEFINITION
ACCESSION   BM616582
VERSION     19600449713643 5', mRNA sequence.
KEYWORDS    EST.
SOURCE      BM616582.1 GI:18914804
ORGANISM    Anopheles gambiae (African malaria mosquito)
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Culicidae; Anophelinae; Anopheles.
            1 (bases 1 to 578)
            Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,
            Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
            Celera Anopheles gambiae EST project
            Unpublished (2002)
            Contact: Holt R.A.
            Celera Genomics
            45 W. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com

```

```

Plate: NU01004AXG row: A column: 21
Seq primer: M13 Reverse.
Location/Qualifiers
1. .578
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449713643"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cdNA1"
/note="Vector: pSport1; Site 1: Sall; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. cDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mr4.org)"

ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 578;
Best Local Similarity 60.8%; Pred. No. 3.6;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 160 CAAGGAACCTGGAGATTGATGGCTTTGACTCTCTAGCCAAATGCCCGCGGTA 219
Db 38 CAACGAGCTGGAGATTGCTCTGGATCAGCCCAACGAGGCTAACGCTGAGGCCCAAGAA 97

QY 220 CTTCAGAGTACCAAGAACTCAAGGCGCGTCAAG 256
Db 98 CATCAAGCGCTACCAGCAGCAGCTGAAGGACGCTCCAG 134

RESULT 11
BM624215 585 bp mRNA linear EST 26-FEB-2002
LOCUS 17000687490727 A.Gam.ad.cdNA1 Anopheles gambiae cDNA clone
DEFINITION 19600449669056 5', mRNA sequence.
ACCESSION BM624215
VERSION 1
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 585)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celeris Anopheles gambiae EST project
Unpublished (2002)
JOURNAL
COMMENT Contact: Holt R.A.
Celeris Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltR@cclera.com
Plate: NU01004AYV row: P column: 14
Seq primer: M13 Reverse.
Location/Qualifiers
1. .585
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449669056"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cdNA1"
/note="Vector: pSport1; Site 1: Sall; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.

FEATURES
source

```

```

cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 585;
Best Local Similarity 60.8%; Pred. No. 3.6;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 160 CAAGGAACCTGGAGATTGATGGCTTTGACTCTCTAGCCAAATGCCCGCGGTA 219
Db 466 CAACGAGCTGGAGATTGCTCTGGATCAGCCCAACGAGGCTAACGCTGAGGCCCAAGAA 525

QY 220 CTTCAGAGTACCAAGAACTCAAGGCGCGTCAAG 256
Db 526 CATCAAGCGCTACCAGCAGCAGCTGAAGGACGCTCCAG 562

RESULT 12
BM628044 607 bp mRNA linear EST 26-FEB-2002
LOCUS 17000687497740 A.Gam.ad.cdNA1 Anopheles gambiae cDNA clone
DEFINITION 19600449652861 5', mRNA sequence.
ACCESSION BM628044
VERSION 1
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 607)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celeris Anopheles gambiae EST project
Unpublished (2002)
JOURNAL
COMMENT Contact: Holt R.A.
Celeris Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltR@cclera.com
Plate: NU010049VE row: M column: 19
Seq primer: M13 Reverse.
Location/Qualifiers
1. .607
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449652861"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cdNA1"
/note="Vector: pSport1; Site 1: Sall; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 607;
Best Local Similarity 60.8%; Pred. No. 3.6;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 160 CAAGGAACCTGGAGATTGATGGCTTTGACTCTCTAGCCAAATGCCCGCGGTA 219
Db 170 CAACGAGCTGGAGATTGCTCTGGATCAGCCCAACGAGGCTAACGCTGAGGCCCAAGAA 229

QY 220 CTTCAGAGTACCAAGAACTCAAGGCGCGTCAAG 256

```

```

Db      230 CATCAAGCGCTACCAGCAGCAGTGAAGGACGTCCAG 266
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17000687493467 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
EST.
BM625696 610 bp mRNA linear EST 26-FEB-2002
19600449628977 5', mRNA sequence.
BM625696
BM625696.1 GI:18925207
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 610)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
CelerA Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
CelerA Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004ABZ row: J column: 15
Seq primer: M13 Reverse.
Location/Qualifiers
FEATURES
1..610
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449628977"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cDNA1"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 610;
Best Local Similarity 60.8%; Pred. No. 3.6;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 160 CAAGGAACCTGGAGATTGGATTGGCTTTGACTCTCTAGCCAAATGCCAGCGGTA 219
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 CATCAGCTGGAGATTGCTTGATCAGCCACACAGGCTAACGCTCAGGCCAGAGAA 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 220 TTTCAGAAGATACAGAACTCAAGGAGCGGTCAG 256
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 424 CATCAAGCGCTACCAGCAGCAGTGAAGGACGTCCAG 460
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
BM578413 615 bp mRNA linear EST 22-FEB-2002
LOCUS 17000687164826 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
DEFINITION 19600449690283 5', mRNA sequence.
ACCESSION BM578413
VERSION BM578413.1 GI:18866880
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Db      230 CATCAAGCGCTACCAGCAGCAGTGAAGGACGTCCAG 266
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17000687493467 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
EST.
BM625696 610 bp mRNA linear EST 26-FEB-2002
19600449628977 5', mRNA sequence.
BM625696
BM625696.1 GI:18925207
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 610)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
CelerA Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
CelerA Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004ABZ row: J column: 15
Seq primer: M13 Reverse.
Location/Qualifiers
FEATURES
1..610
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449628977"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cDNA1"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 610;
Best Local Similarity 60.8%; Pred. No. 3.6;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 160 CAAGGAACCTGGAGATTGGATTGGCTTTGACTCTCTAGCCAAATGCCAGCGGTA 219
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 CATCAGCTGGAGATTGCTTGATCAGCCACACAGGCTAACGCTCAGGCCAGAGAA 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 220 TTTCAGAAGATACAGAACTCAAGGAGCGGTCAG 256
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 424 CATCAAGCGCTACCAGCAGCAGTGAAGGACGTCCAG 460
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
BM618626 630 bp mRNA linear EST 25-FEB-2002
LOCUS 17000687439284 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
DEFINITION 19600449662409 5', mRNA sequence.
ACCESSION BM618626
VERSION BM618626.1 GI:18917044
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 630)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
CelerA Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
CelerA Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004185 row: K column: 15
Seq primer: M13 Reverse.

```

